DOCUMENT RESUME

ED 324 373 TM 015 650

AUTHOR Harrison, Ferrin; Katti, S. K.
TITLE Optimal Design in and Hazards of Linearization of Langmuir's Nonlinear Model.
PUB DATE Jun 90
NOTE 28p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Equations (Mathematics); *Estimation (Mathematics); *Least Squares Statistics; *Mathematical Models; Regression (Statistics); Simulation; *Statistical Analysis
IDENTIFIERS *Langmuir Model; Linear Models; Nonlinear Models; Optimal Scaling

ABSTRACT

Langmuir's model is studied for the situation where epsilon is independently and identically normally distributed. The "Y/x" versus "Y" plot had a 90% mid-range that did not contain the true curve in a vast portion of the range of "x". The "1/Y" versus "1/chi" plot had undefined expected values, and this problem worsens as sample size increases. The use of non-linear least squares is recommended. In non-linear regression, it is demonstrated that a design that gives at least a local minimum of the generalized variance of the parameter estimators is one where half the observations are taken at the maximal value of "x" (termed "x-theta") and half of the epsilons equal half their maximal value. The use of such an extreme design is optimal in simple linear regression; however, it is curious that this design is optimal in non-linear Langmuir's model. Six graphs, five data tables, and numerous equations are provided. (Author/SLD)

OPTIMAL DESIGN IN AND HAZARDS OF LINEARIZATION OF LANGMUIR'S NONLINEAR MODEL

FERRIN HARRISON
by
and
S. K. KATTI

June 1990

Abstract

We study Langmuir's model, $Y = \frac{a\mathbf{x}}{1+b\mathbf{x}} + \epsilon$, where ϵ is independently and identically normally distributed, $N(0, \sigma^2)$. The $Y/\mathbf{x}$ versus Y plot had a 90% midrange which did not contain the true curve in a vast portion of the range of $\mathbf{x}$. The $1/Y$ versus $1/x$ plot has undefined expected values, and gets worse as the sample size increases. We recommend the use of nonlinear least squares.

In nonlinear regression, we prove that a design which gives at least a local minimum of the generalized variance of the parameter estimators is one where half the observations are at the maximum value of $\mathbf{x}$, called $\mathbf{x}_\infty$, and the remainder at $\mathbf{x} = \frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty}$. The use of such an extreme design is optimal in simple linear regression. It is curious that it is optimal in nonlinear Langmuir's model.

Key Words:

Langmuir's model	.90 MidRange
(statistical) simulation	analytic methods

I. Introduction

A. BASIC MATERIAL: The Langmuir model or isotherm is defined by

$$Y = \frac{a\mathbf{x}}{1+b\mathbf{x}} + \epsilon$$

where $\mathbf{x}$ is the control variable, free from error. The response variable Y, is measured with error ϵ, which is *not* a function of $\mathbf{x}$. We assume that the errors ϵ_i, $i = 1, \ldots, n$ are i.i.d. $N(0, \sigma^2)$ both for the sake of simplicity and to go along with the literature. Our objective is to estimate a and b.

Chemical engineers and biologists who wish to estimate a and b commonly avoid minimizing this sum of squares:

$$\chi^2 = \sum_{i=1}^{n} \left(Y_i - \frac{a\mathbf{x}_i}{1+b\mathbf{x}_i} \right)^2 \tag{1}$$

Instead, they minimize some other sum of squares, such as the following:

Linearization I:

$$\sum_{i=1}^{n} \left(Y_i - a\mathbf{x}_i + b\mathbf{x}_i Y_i \right)^2 \tag{2}$$

Linearization II, Scatchard:

$$\sum_{i=1}^{n} \left(\frac{Y_i}{\mathbf{x}_i} - a + bY_i \right)^2 \quad \text{or plot } \frac{Y}{\mathbf{x}} \text{ versus } Y \tag{3}$$

Linearization III, Lineweaver–Burk:

$$\sum_{i=1}^{n} \left(\frac{1}{Y_i} - \frac{b}{a} - \frac{1}{a\mathbf{x}_i} \right)^2 \quad \text{or plot } \frac{1}{Y} \text{ versus } \frac{1}{\mathbf{x}} \tag{4}$$

Each transformation to linearize the model, or linearization, is obtained by modifying $Y = a\mathbf{x}/(1+b\mathbf{x})$ without regard to ϵ. Presumably, nonlinear

models are linearized not because of the structure of the error, but to avoid difficulties in numerically obtaining nonlinear least square estimates, and permit graphing a straight line. The question which we raise is: *what is the effect of the error term on such a linearization and on the inference made?*

We investigate these methods of estimating the parameters, for the usual design of the control variable, $\mathbf{x}$. Our conclusion, along with that of Colquhoun[1] is that the linearization does not work! On the other hand, nonlinear estimation poses no particular problem when one has access to a computer. We also develop designs of $\mathbf{x}$ values to obtain better parameter estimates, under the assumption that the Langmuir's model is true.

B. DERIVATION OF THE MODEL: Langmuir[2] derived his isotherms in 1916. We use chemistry notation to derive the model. A compound of components A and S is written AS. The concentration of a compound A is written $[A]$, using the brackets to denote the concentration measure.

For one mobile phase compound A, and stationary phase compound S, the reaction $A + S \rightleftharpoons AS$ has the reaction equilibrium constant

$$k = \frac{[AS]}{[A][S]}.$$

By the conservation of mass, the number of components of S is some constant c, where $[S] + [AS] = c$. So

$$k = \frac{[AS]}{[A](c - [AS])},$$

and $ck[A] = [AS] + k[A][AS]$, and

$$[AS] = \frac{ck[A]}{1 + k[A]}.$$

To translate the model into the usual notation of statistics, define $a = ck$, $b = k$, $Y = [AS]$, and $\mathbf{x} = [A]$, so $\mathcal{E}(Y) = \frac{a\mathbf{x}}{1+b\mathbf{x}}$. The model can be derived for Michaelis-Menten reactions similarly, shown in many textbooks.

Doug Doren, a chemist at the University of Delaware, knew that one can optimize over parameter estimation by taking equally many observations as far apart as possible after linearization. Based on the derivation of the Michaelis-Menten, we suspect that some people have guessed our optimal design. It seems that no one has shown its optimality using analytical methods before us. So our claim to fame is in proving something which the best minds have only suspected.

C. NONLINEAR ESTIMATION: In general, one finds the simultaneous zero of the partial derivatives of χ^2 to minimize χ^2 in equation (1): $\frac{\partial \chi^2}{\partial a^*} = 0$ and $\frac{\partial \chi^2}{\partial b^*} = 0$.

In our problem, we found a transformation of these equations helpful. Since the model is linear in parameter a, we can solve in closed form to obtain a quantity $h(b)$ such that $\hat{a} = h(\hat{b})$:

$$h(b) = \frac{\sum_{i=1}^{n} Y_i \cdot \frac{\mathbf{x}_i}{1+b\mathbf{x}_i}}{\sum_{i=1}^{n} \left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^2}$$

Next, we transform the second equation, to an equation of the form

$$r \cdot \frac{\partial \chi^2}{\partial a^*} + \frac{\partial \chi^2}{\partial b^*} = 0.$$

We take the approach of taking the derivative of $\chi^2[h(b^*),\ b^*]$ with respect to b^* to obtain:

$$\frac{\mathrm{d}h(b^*)}{\mathrm{d}b^*} \frac{\partial \chi^2}{\partial h(b^*)} + \frac{\partial \chi^2}{\partial b^*} = 0.$$

Define $\mathbf{z}_i = \frac{\mathbf{x}_i}{1+b\mathbf{x}_i}$ to simplify the expression. So, the derivative of $h(b)$ with respect to b is

$$h'(b) = \frac{\mathrm{d}h(b)}{\mathrm{d}b} = \frac{-\sum_{i=1}^{n} Y_i \mathbf{z}_i^2 \sum_{i=1}^{n} \mathbf{z}_i^2 + 2\sum_{i=1}^{n} \mathbf{z}_i^3 \sum_{i=1}^{n} Y_i \mathbf{z}_i}{\left[\sum_{i=1}^{n} \mathbf{z}_i^2\right]^2}$$

We complete the evaluation of the formula to obtain the formula used on the computer:

$$g(b) = -\sum_{i=1}^{n} \left[Y_i - \frac{h(b)\mathbf{x}_i}{1 + b\mathbf{x}_i} \right] \frac{\mathbf{x}_i}{1 + b\mathbf{x}_i} \left[h'(b) - \frac{h(b)\mathbf{x}_i}{1 + b\mathbf{x}_i} \right] = 0$$

We solve this equation for $\hat{b}$ to obtain $\hat{a} = h(\hat{b})$, without checking $\frac{\partial \chi^2}{\partial a^*} = 0$, since we have already solved that problem analytically.

We use bisection, or binary search to find b^*, where b^* is such that $g(b^*) \simeq 0$. We define the sign function as follows:

$$\text{sgn}(f) = \begin{cases} 1 & \text{if } g > 0 \\ 0 & \text{if } g = 0 \\ -1 & \text{if } g < 0 \end{cases}$$

We first find values b_1, b_3 such that the signs are different, $\text{sgn}[g(b_1)] \neq \text{sgn}[g(b_3)]$. Next, assign $b_2 \leftarrow \frac{b_1 + b_3}{2}$. If $\text{sgn}[g(b_1)] = \text{sgn}[g(b_2)]$, we assign $b_1 \leftarrow b_2$. Otherwise, if the solution is not in hand, we assign $b_3 \leftarrow b_2$. The process is repeated till we know the least squares b to enough digits. Bisection gave reasonable answers for 32 000 simulations. Hence this method is recommended to obtain the nonlinear least squares for Langmuir's model. We do not need the general purpose nonlinear estimation schemes such as the maximum descent, along with the transformations of Habibullah[3]. For the two and more solute extensions of Langmuir's model, general purpose numerical minimizations might be necessary.

II. Simulation Based Results 1

A. LINEAR III WORSE AS SAMPLE INCREASES: For each simulation in this section, we use the design $(1/n,\ 2/n,\ \ldots,\ n/n)$. For each $\mathbf{x}_i$, we have an expected value of Y_{ij}, that is $\frac{a\mathbf{x}_i}{1+b\mathbf{x}_i}$, for $\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_n$. So

$$Y_{ij} = \frac{a\mathbf{x}_i}{1 + b\mathbf{x}_i} + \epsilon_{ij},$$

for $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, 2000$. The ϵ_{ij} come from a normal random number generator, with mean zero and standard deviation 0.2. For each sample $(Y_{ij}, \mathbf{x}_{ij})$, $i = 1, \ldots, n$, we estimate (a, b). So we have 2000 estimates of (a, b) for each case considered. We used $a = 25$ and $b = 10$, the usual parameter estimates (rounded to integer values) given by Anita Katti, from her previous work at Oak Ridge National Labs.

Note that at $\mathbf{x} = 0$, $\mathcal{E}(Y) = 0$. So, the relative error is infinite. At $\mathbf{x} = \infty$, the relative error is 8%. Since the expected value of Y_i is proportional to a, and we are using least squares, root n asymptotics apply. So we expect similar results for other parameterizations with our values of $\sigma/(n^{1/2}a)$. We chose a combination of standard deviation and sample size to illustrate our results.

The "range" of estimates $\hat{a}$ is the largest of the 2000 simulated estimates less the smallest. The $\Pr(\hat{b} < 0)$ is the probability observed in the simulation.

Table 1	Estimates $\hat{a}$, with $a = 25$, $\sigma = .2$.			
size:	10	40	160	640
III range:	90	73	95 338	131 000
NL range:	81	25	12	7
III min:	7	$^-$16	$^-$12	$^-$46 460
NL min:	12	17	20	22
$\Pr(\hat{b} < 0)$:	0	2%	31.8%	39.7%

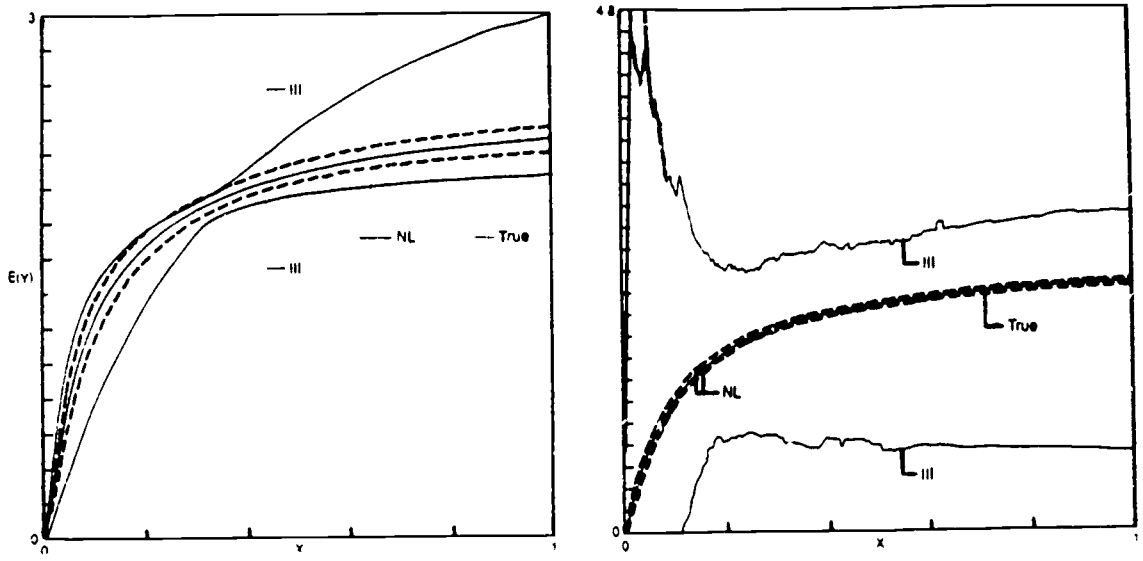
Table 1 shows the behavior of methods III and NL for increasing sample sizes, picking $\mathbf{x} = \{\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n}{n}\}$.

Since the density of Y is continuous and nonzero at zero, and linearization III contains $\frac{1}{Y}$, the theoretical moments are infinite or undefined. This shows up as the vast range 131 000 for $\hat{a}$ when method III is used, while the range for the nonlinear decreases to 7. Thus, the probabilities that gave reasonable midranges for method III and small sample sizes are less helpful for larger sample sizes.

Figure 1 contains a graph of $\hat{Y} = \frac{\hat{a}\mathbf{x}}{1+\hat{b}\mathbf{x}}$. We have 2000 graphs. To reduce volume, we take a point $\mathbf{x}$ on the $\mathbf{x}$-axis. We compute all of the

2000 $\hat{Y}$, and rank them. We throw away the lowest and the highest 5% or 100 of the $\hat{Y}$-values to obtain an interval which contains the remaining 90%. We plot the 101st and 1900th points. We call them the .90 midrange. Linearization III has **much larger** midranges than nonlinear.

Linear III versus Nonlinear .90 MidRanges



A. Sample Size n=40 B. Sample Size n=640

Fig. 1.

Figure 1 is a comparison between linearization "III" and nonlinear, "NL" least squares. The sample sizes are 40 and 640. The true curve is $\mathcal{E}(Y) = 25\mathbf{x}/(1 + 10\mathbf{x})$, and is labeled "true." The midrange curves are labeled "III" for estimation using linearization "III". The midrange curves for nonlinear least squares are labeled "NL". The standard deviation, $\sigma = \mathrm{Var}^{1/2}(\epsilon)$, is 0.2 in the 2000 simulated samples used.

For the cases of n=40 and n=640 values of $\mathbf{x}$, located at $\frac{i}{n}$, $i = 1, \ldots, n$, we see that the nonlinear least squares midrange is nearly symmetric about the true curve and closer to the true curve than the linearization III midranges for the vast majority of values of $\mathbf{x}$. Linearization III given in Figure 3, part B with sample size 640 is distinctly worse than the curve in part A for a sample size of only 40. This is due to the reciprocal. For n=640, there are more observations near $\mathbf{x}$=0 than for n=40.

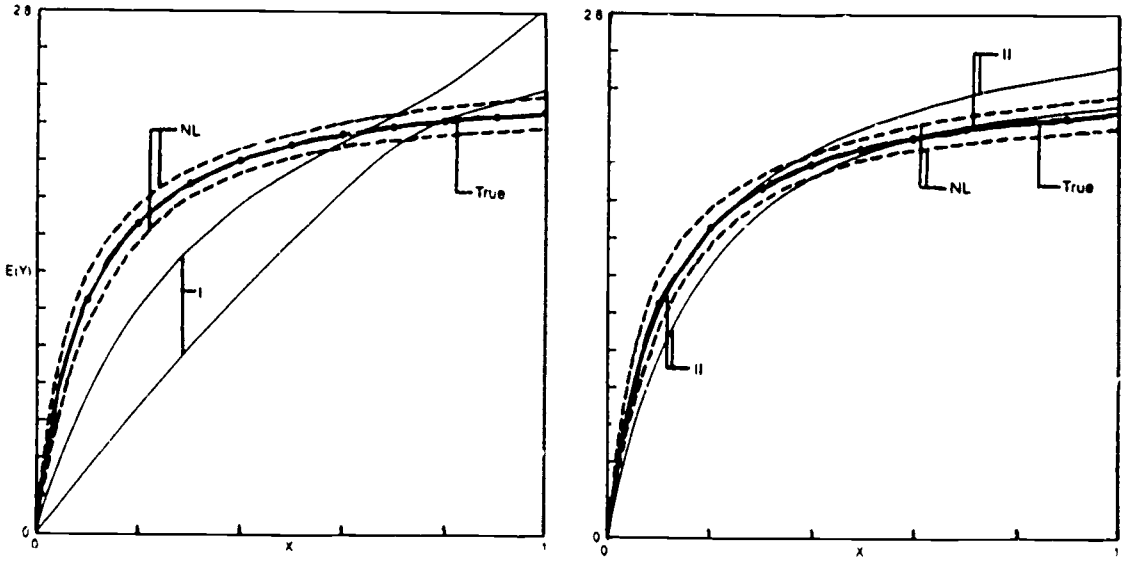
At sample size 640, about 40% of the estimates of b are negative for linearization III, given in table 1. These negative estimates lack physical meaning, and yield the negative midrange curve values for $\mathbf{x} < .1$, for linearization III.

The loss of accuracy with increasing data, shown in table 1 and figure 3, is not acceptable. So, we set aside linearization III.

B. FURTHER COMPARISONS BY SIMULATION: Next, we compare estimates obtained through least squares on linearizations I and II with nonlinear least squares. We chose the following thirty values of $\mathbf{x}$.

$$(\mathbf{x}_i) = (.1,\ .1,\ .1,\ .2,\ .2,\ .2,\ .3,\ \ldots,\ .9,\ 1,\ 1,\ 1).$$

Linear I, II versus Nonlinear .90 MidRanges



A. Linear I versus Nonlinear B. Linear II versus Nonlinear

Fig. 2.

Figure 2 contains graphs of the midranges for the methods Linearization I, Linearization II, and NonLinear (NL) estimation. Circles are drawn on the *true* curve, $\mathcal{E}(Y) = 25\mathbf{x}/(1 + 10\mathbf{x})$, to indicate the locations of the observations. The midrange curves for nonlinear regression are in dashes.

We use Figure 2 to compare linearizations "I" and "II" with nonlinear least squares. The standard deviation is 0.2, and the design is $(\mathbf{x}_i) = (.1,\ .1,\ .1,\ .2,\ .2,\ .2,\ .3,\ \ldots\ 1)$. The midrange curves for nonlinear least squares are labeled "NL", and consistently contain the true value, quite unlike the midranges for the linearizations.

Table 2 For Parameter estimates $\hat{a}$, for $a = 25$
Design $(\mathbf{x}) = (.1,\ .1,\ .1,\ .2,\ \ldots\ ,\ 1)$, $\sigma = 0.2$

	Linear I	Linear II	Nonlinear
range:	15.508	21.182	34.8
Cor$(\hat{a},\ \hat{b})$:	.99997	.9989	.99435
mean:	6.467	17.885	25.559

To get a better assessment of the situation, we look at the estimated co-relations in Table 2. We use the following mathematization of co-relation, for estimators $\hat{\theta}_1,\ \hat{\theta}_2$, with parameters $\theta_1,\ \theta_2$:

$$\text{Co} - \text{relation}(\hat{\theta}_1,\ \hat{\theta}_2) = \frac{\mathcal{E}(\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2)}{\left[\mathcal{E}(\hat{\theta}_1 - \theta_1)^2 \mathcal{E}(\hat{\theta}_2 - \theta_2)^2\right]^{1/2}}.$$

To estimate the co-relation, we substitute means for expectations. The estimated co-relations are .994 and up, so the parameter estimates are highly co-related, making accurate estimation difficult. These co-relations and the nonelliptic squared error contour graphs of Colquhoun[1] imply that finding the minimum using the maximum descent method can be difficult. Habibullah[3] transforms such contours to approximately circular contours by transforming the parameters. He transformed concentric banana shaped contours, which apparently were not asymptotically elliptic, to roughly circular contours. In one of the problems, the reduction was from 13010 iterations for the usual maximum descent to 19 iterations after the transform, and from 9425 iterations to 41 in another. Eastham et al.[4] found transformation of the parameters quite useful in finding maximum likelihood estimators for the three parameter lognormal distribution. We expect the trick of transforming the parameters to permit the use of numerical min-

inizing even when the two and more solute extensions of the Langmuir's models do not permit the simplicity of using bisection as it was done here.

In view of the regions in which the midrange *does not include the true curve*, and Table 2, we discard methods I and II.

C. EVIDENCE FOR ASYMPTOTIC NORMALITY: Next we look at nonlinear regression with uniform design, in table 3. The mean goes to the true value 10 as $n \to \infty$. The mean squared error of the estimator goes to zero at the proper rate of $\frac{1}{n}$. Thus, when multiplied by $\frac{n}{10}$, the mean squared error is as constant as can be expected from simulation.

Table 3 Moments of $\hat{b}$ by simulation
For $b = 10$, nonlinear

size:	10	40	160	640
mean:	10.87	10.18	10.06	10.02
$\frac{n}{10}$MSE :	15.71	10.71	10.61	11.05
skew:	1.55	.57	.31	.17
kurt:	7.84	3.56	3.32	3.32

Our skewness is the mean of cubed error from the true value of the parameter, and kurtosis is the mean of error to the fourth power. We make the skewness and kurtosis dimensionless, independent of scale by dividing by the mean squared error to the powers $\frac{3}{2}$ and 2, respectively. We show these moments (estimated from 2000 simulations) in table 3.

The skewness is going to zero, and the kurtosis to three, which is the value of the kurtosis for the standard normal distribution. This implies rapid convergence of the distribution of parameter estimates to the normal distribution.

So, one can expect the asymptotic normal approximation to the finite sample distribution of $(\hat{a}.\ \hat{b})$ to be a good approximation. The asymptotic formula is derived in Chapter III.

III. Analytic Results

A. ASYMPTOTIC PARAMETER ESTIMATE VARIANCES: Define $(a,\ b)$ as the constant, true values of the parameters, $(a^*,\ b^*)$ as variables, and $(\hat{a},\ \hat{b})$ as parameter estimates. Also, $y_i = a\mathbf{x}_i/(1+b\mathbf{x}_i) + \epsilon_i$. The control variable is $\mathbf{x}$, and the response variable is y.

We recall our sum of squares,

$$Q = \sum_{i=1}^{n} \left(y_i - \frac{a^*\mathbf{x}_i}{1+b^*\mathbf{x}_i} \right)^2$$

$$\frac{\partial Q}{\partial a^*} = 0 \qquad \text{so}$$

$$0 = \sum_{i=1}^{n} \epsilon_i \frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} + \mu_i \frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} - \hat{a}\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2$$

$$\frac{\partial Q}{\partial b^*} = 0 \qquad \text{so}$$

$$0 = \sum_{i=1}^{n} \epsilon_i \hat{a}\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2 + \mu_i \hat{a}\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2 - \hat{a}^2\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^3$$

We assume that we will use reasonable sequences of designs of experiment. We assume that for reasonable designs of experiment, $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b. We take a large enough sample that $\Delta a = \hat{a} - a$, and $\Delta b = \hat{b} - b$ are infinitesimals with high probability. Define:

$$f_1 = \sum_{i=1}^{n} \epsilon_i \frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} + \mu_i \frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} - \hat{a}\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2$$

and

$$f_2 = \sum_{i=1}^{n} \epsilon_i \left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2 + \mu_i \left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^2 - (a+\Delta a)\left(\frac{\mathbf{x}_i}{1+\hat{b}\mathbf{x}_i} \right)^3$$

The model is linear in a, so the functions are already expanded in Δa. We expand by Taylor series in Δb only, about zero.

$$
\begin{aligned}
& f_1(\epsilon;\ \Delta a;\ \Delta b) \\
& \quad \simeq\ f_1(\epsilon;\ \Delta a;\ 0) + \Delta b \frac{\partial f_1}{\partial \Delta b}(\epsilon;\ \Delta a;\ 0) \\
& \quad \simeq\ \sum_{i=1}^{n} \epsilon_i \frac{\mathbf{x}_i}{1+b\mathbf{x}_i} + \mu_i \frac{\mathbf{x}_i}{1+b\mathbf{x}_i} - (\hat{a})\left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^2 - \Delta b \epsilon_i \left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^2 \\
& \quad + \sum_{i=1}^{n} -\Delta b \mu_i \left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^2 2\Delta b(\hat{a})\left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^3 \\
& \quad =\ \sum_{i=1}^{n} \epsilon_i \mu_i a^{-1} - \Delta a \mu_i^2 a^{-2} - \Delta b \epsilon_i \mu_i^2 a^{-2} + \Delta b \mu_i^3 a^{-2},
\end{aligned}
$$

correct to infinitesimal of the first order.

Define $\overline{\mu^k} = \frac{1}{n}\sum_{i=1}^{n} \mu_i^k$, and similarly for $\overline{\epsilon\mu^k} = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \mu_i^k$.

$$
\frac{1}{n} f_1(\epsilon;\ \Delta a;\ \Delta b) = \overline{\epsilon\mu} a^{-1} - \Delta a \overline{\mu^2} a^{-2} - \Delta b \overline{\epsilon\mu^2} a^{-2} + \Delta b \overline{\mu^3} a^{-2},
$$

Next, we multiply by $\frac{a^2}{n}$ to obtain

$$
0 \simeq a\overline{\epsilon\mu} - \Delta a \overline{\mu^2} + \Delta b \overline{\mu^3} - \Delta b \overline{\epsilon\mu^2}.
$$

Now we consider the term $\overline{\epsilon\mu^2}$, in the Δb term. Since the ϵ_i's are i.i.d. $N(0,\ \sigma^2)$,

$$
\begin{aligned}
\mathrm{Var}(\overline{\epsilon\mu^2}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \epsilon_i \mu_i^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(\epsilon_i)\mu_i^4 \\
&= \frac{1}{n}\sigma^2 \overline{\mu^4}
\end{aligned}
$$

So, $\overline{\epsilon\mu} \sim N(0,\ \frac{1}{n}\sigma^2\overline{\mu^4})$, and the term $-\Delta b\overline{\epsilon\mu^2}$ is an infinitesimal of higher order than Δb. Hence, ignoring the $-\Delta b\overline{\epsilon\mu^2}$ term in comparison with Δa and Δb:

$$a(\overline{\epsilon\mu}) \simeq [\overline{\mu^2} \quad -\overline{\mu^3}] \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix}$$

Next, correct to infinitesimal of the first order,

$$\begin{aligned} f_2(\epsilon;\ \Delta a;\ \Delta b) &= f_2(\epsilon;\ \Delta a;\ 0) + \Delta b \frac{\partial f_2}{\partial \Delta b}(\epsilon;\ \Delta a;\ 0) \\ &= \sum_{i=1}^{n} \epsilon_i \mu_i^2 a^{-2} + \mu_i \mu_i^2 a^{-2} - (a + \Delta a)\mu_i^3 a^{-3} \\ &+ \sum_{i=1}^{n} -2\Delta b \epsilon_i \mu_i^3 a^{-3} - 2\mu_i \Delta b \mu_i^4 a^{-3} + 3\Delta b \mu_i^4 a^{-3} \end{aligned}$$

We multiply by $\frac{1}{n}a^3$ to obtain

$$0 = a(\overline{\epsilon\mu^2}) - \Delta a\overline{\mu^3} - 2\Delta b(\overline{\epsilon\mu^3}) + \Delta b\overline{\mu^4}$$

Similarly, we drop the $\Delta b\overline{\epsilon\mu^3}$ term. yielding $\mathrm{Var}(\overline{\epsilon\mu^3}) = \frac{1}{n}\sigma^2\overline{\mu^6}$. Now we discard the $\Delta b\overline{\epsilon\mu^3}$ term, leaving

$$a(\overline{\epsilon\mu^2}) \simeq [\overline{\mu^3} \quad -\overline{\mu^4}] \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix}.$$

Hence,

$$\begin{bmatrix} a\overline{\epsilon\mu} \\ a\overline{\epsilon\mu^2} \end{bmatrix} \simeq \begin{bmatrix} \overline{\mu^2} & -\overline{\mu^3} \\ \overline{\mu^3} & -\overline{\mu^4} \end{bmatrix} \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix} = \mathrm{M} \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix},$$

and

$$\begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix} \simeq a\mathrm{M}^{-1} \begin{bmatrix} \overline{\epsilon\mu} \\ \overline{\epsilon\mu^2} \end{bmatrix}.$$

So the covariance can be found as

$$\left[\mathrm{Cov}(\hat{a},\ \hat{b})\right] = a^2\mathbf{M}^{-1}\begin{bmatrix}\mathrm{Var}(\overline{\epsilon\mu}) & \mathrm{Cov}(\overline{\epsilon\mu},\ \overline{\epsilon\mu^2})\\ \mathrm{Cov}(\overline{\epsilon\mu},\ \overline{\epsilon\mu^2}) & \mathrm{Var}(\overline{\epsilon\mu^2})\end{bmatrix}\mathbf{M}^{-1T}$$

where

$$\mathbf{M}^{-1} = \frac{1}{\overline{\mu^2}\,\overline{\mu^4} - \overline{\mu^3}^2}\begin{bmatrix}-\overline{\mu^4} & \overline{\mu^3}\\ -\overline{\mu^3} & \overline{\mu^2}\end{bmatrix}.$$

The need for $\mathbf{M}$ to be nonsingular makes it necessary that $\overline{\mu^2}\cdot\overline{\mu^4} - \overline{\mu^3}^2$ is not 0. A sufficient condition to make $\mathbf{M}$ nonsingular is that there be two different and nonzero values of $\mathbf{x}$ in the design, shown in other writings. Next, recalling the independence of the ϵ's,

$$\left[\mathrm{Cov}(\overline{\epsilon\mu},\ \overline{\epsilon\mu^2})\right] = \begin{bmatrix}\frac{\sigma^2}{n}\overline{\mu^2} & \frac{\sigma^2}{n}\overline{\mu^3}\\ \frac{\sigma^2}{n}\overline{\mu^3} & \frac{\sigma^2}{n}\overline{\mu^4}\end{bmatrix}$$

So

$$\begin{aligned}&\left[\mathrm{Cov}(\hat{a},\ \hat{b})\right]\\ &= a^2\mathbf{M}^{-1}\left[\mathrm{Cov}(\overline{\epsilon\mu},\ \overline{\epsilon\mu^2})\right]\mathbf{M}^{-1T}\\ &= \frac{a^2\sigma^2}{n\left(\overline{\mu^2}\cdot\overline{\mu^4} - \overline{\mu^3}^2\right)^2}\begin{bmatrix}-\overline{\mu^4} & \overline{\mu^3}\\ -\overline{\mu^3} & \overline{\mu^2}\end{bmatrix}\begin{bmatrix}\overline{\mu^2} & \overline{\mu^3}\\ \overline{\mu^3} & \overline{\mu^4}\end{bmatrix}\begin{bmatrix}-\overline{\mu^4} & -\overline{\mu^3}\\ \overline{\mu^3} & \overline{\mu^2}\end{bmatrix}\\ &= \frac{a^2\sigma^2}{n\left(\overline{\mu^2}\cdot\overline{\mu^4} - \overline{\mu^3}^2\right)^2}\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}-\overline{\mu^4} & -\overline{\mu^3}\\ \overline{\mu^3} & \overline{\mu^2}\end{bmatrix}\\ &= \frac{a^2\sigma^2}{n\left(\overline{\mu^2}\cdot\overline{\mu^4} - \overline{\mu^3}^2\right)^2}\begin{bmatrix}\overline{\mu^4} & \overline{\mu^3}\\ \overline{\mu^3} & \overline{\mu^2}\end{bmatrix}.\end{aligned}$$

It is consoling to note that this matches with the formula in Oliver[5], up to a clerical error in Oliver[5]. The error can be removed by substituting

the definitions

$$B = \sum \frac{x^3}{(K+x)^3}, \qquad C = \sum \frac{x^4}{(K+x)^4},$$

for the definitions

$$B = \sum \frac{x^2}{(K+x)^3}, \qquad C = \sum \frac{x^2}{(K+x)^4}.$$

Next, we check for concordance or discordance between the asymptotic results and simulation results. We standardize the mean squared errors by multiplying by the sample size, so the variances will not go to zero as n goes to infinity. For the design $(\mathbf{x}) = (.1, .2, .3, \ldots, 1)$, and replicates of this design, with parameters $a = 25$, $b = 10$, and $\sigma = .2$, the matrix of standardized covariances is as follows.

$$\begin{bmatrix} n\mathrm{Var}(\hat{a}) & n\mathrm{Cov}(\hat{a}, \hat{b}) \\ n\mathrm{Cov}(\hat{a}, \hat{b}) & n\mathrm{Var}(\hat{b}) \end{bmatrix} \doteq \begin{bmatrix} 488.699 & 232.302 \\ 232.302 & 111.829 \end{bmatrix}$$

Table 4 Comparison of Finite and Asymptotic Moments

	Simulation				Asymptotic
n	10	40	160	320	$[\mathrm{Cov}(\hat{a}, \hat{b})]$
$n\mathrm{MSE}(\hat{a})$	605.4	508.5	502.2	497.3	488.7
$n\mathrm{Cov}(\hat{a}, \hat{b})$	325.8	250.1	237.4	235.5	232.3
$n\mathrm{MSE}(\hat{b})$	141.8	117.6	115.2	114.2	111.8

We compare the standardized mean squared errors and covariances obtained from asymptotics with coefficients from simulation, using the same designs. The $\overline{\mu^n}$'s and variance-covariance matrix above are the same for any number of replications of the same design.

With increasing sample size, the simulation based estimates approach the asymptotic based approximation from above.

B. OPTIMALITY CRITERION AND ALGORITHM: Next, we define the standardized, generalized variance as the determinant of the matrix of standardized covariances. We define the optimal design as the design minimizing the standardized, generalized variance. We use the asymptotic formulae as approximations to the true values. So the standardized, generalized variance is asymptotically:

$$\begin{aligned}\mathrm{SGVar}(\hat{a},\ \hat{b}) &= \begin{vmatrix} n\mathrm{Var}(\hat{a}) & n\mathrm{Cov}(\hat{a},\ \hat{b}) \\ n\mathrm{Cov}(\hat{a},\ \hat{b}) & n\mathrm{Var}(\hat{b}) \end{vmatrix} \\ &= \frac{\sigma^4 a^4}{\overline{\mu^2}\cdot\overline{\mu^4} - \overline{\mu^3}^2}.\end{aligned}$$

Minimizing this quantity is equivalent to maximizing its denominator divided by a^6 with respect to $\mathbf{x}_i$,

$$G^* = \left[\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^2\right]\left[\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^4\right] - \left[\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}\right)^3\right]^2.$$

Define

$$\nu_i = \frac{1+b\mathbf{x}_\infty}{\mathbf{x}_\infty}\frac{\mathbf{x}_i}{1+b\mathbf{x}_i},$$

so ν takes exactly the range [0, 1].

Maximizing G^* with respect to $(\mathbf{x})$ is equivalent to maximizing with respect to the vector (ν) :

$$G = \left(\frac{1+b\mathbf{x}_\infty}{\mathbf{x}_\infty}\right)^6 G^* = \left[\sum_{i=1}^{n}\nu_i^2\right]\left[\sum_{i=1}^{n}\nu_i^4\right] - \left[\sum_{i=1}^{n}\nu_i^3\right]^2.$$

Note that this G is not a function of a nor σ^2. Hence the optimal design is the same for all a and σ^2. Given (ν_i), the experimenter should recognize an approximate value of b and get $(\mathbf{x}_i)$ by solving $\nu_i = \frac{1+b\mathbf{x}_\infty}{\mathbf{x}_\infty}\frac{\mathbf{x}_i}{1+b\mathbf{x}_i}$. So we work with these ν_i, $\nu_i \in [0,\ 1]$ to solve the problem in generality and transform to real units when finished.

If we multiply each ν_i by some constant $k > 1$, then the new value of G is the old value multiplied by k^6. Since each $\mathbf{x}_i$ is in $[0, \infty)$, each ν_i is between zero and one, and the optimal design contains at least one $\nu_i = 1$. We set $\nu_1 = 1$.

Our algorithm is as follows. For any $n-1$ values, we find the optimal nth value to add. When we have enough observations, we find the best value of ν_2 based on the other $\nu_i's$. We continue for ν_i, $i = 2, 3, \ldots, n$, and repeat until the algorithm converges.

We define

$$S(l, m) = \sum_{i \in \{1, n\} \sim m} \nu_i^l = \left(\sum_{i=1}^{m-1} \nu_i^l \right) + \left(\sum_{i=m+1}^{n} \nu_i^l \right)$$

To take partial derivatives with respect to ν_m, we express G as

$$\begin{aligned} G &= \left[S(2, m) + \nu_m^2\right] \left[S(4, m) + \nu_m^4\right] - \left[S(3, m) + \nu_m^3\right]^2 \\ &= S(2, m)S(4, m) - S^2(3, m) \\ &\quad + S(2, m)\nu_m^4 + S(4, m)\nu_m^2 - 2S(3, m)\nu_m^3 \end{aligned}$$

$$\begin{aligned} \frac{\partial G}{\partial \nu_m} = 0 &= 4S(2, m)\nu_m^3 - 6S(3, m)\nu_m^2 + 2S(4, m)\nu_m \\ &= 2\nu_m \left[2S(2, m)\nu_m^2 - 3S(3, m)\nu_m + S(4, m)\right] \end{aligned}$$

with solutions

$$\nu_m = 0, \ \frac{3S(3, m) \pm \left[9S^2(3, m) - 8S(2, m)S(4, m)\right]^{\frac{1}{2}}}{4S(2, m)}.$$

The double partial derivative should be negative.

$$\frac{\partial^2 G}{\partial \nu_m^2} = 12S(2, m)\nu_m^2 - 12S(3, m)\nu_m + 2S(4, m).$$

For $\nu_m = 0$, the double partial is $2S(4,\ m) > 0$ so we can discard $\nu_m = 0$. For the other two roots, define the discriminant as

$$d = 9S^2(3,\ m) - 8S(2,\ m)S(4,\ m).$$

$$\frac{\partial^2 G}{\partial \nu_m^2} = \frac{12S(2,\ m)}{16S^2(2,\ m)}[3S(3,\ m) \pm d^{\frac{1}{2}}]^2 - \frac{12S(3,\ m)}{4S(2,\ m)}[3S(3,\ m) \pm d^{\frac{1}{2}}] + 2S(4,\ m).$$

Next,

$$2S(2,\ m)\frac{\partial^2 G}{\partial \nu_m^2} = 9S^2(3,\ m) - 8S(2,\ m)S(4,\ m) \pm 3S(3,\ m)d^{\frac{1}{2}}.$$

The smaller double derivative could be negative or zero, if the discriminant is zero:

$$\frac{\partial^2 G}{\partial \nu_m^2} = \frac{d \pm 3S(3,\ m)d^{\frac{1}{2}}}{2S(2,\ m)}.$$

The interest in this value of ν_m is not based on knowing that it is a relative maximum, but on knowing that *it is the only interior point which could be a relative maximum.*

Next, we used APL (A Programming Language, source code available on request,) to compute the sequence of conditionally optimal observations. Table 5 contains the results. The code was set up to choose $\nu_i = .5$ when there was a choice, that is, when the value of G is the same for either choice. All aspects of this table inspire the lemma to follow. Firstly, after the initial three, the designs added are alternating between .5 and 1. There is a choice for odd sample sizes, but no choice otherwise.

Table 5

Table of Steps in the Optimal Design Algorithm

	Values of G		
Observation	$\nu = .5$	$\nu = 1$	ν chosen
1	—	—	1
2	.0625	0	.5
3	.125	.125	.5
4	.1875	.25	1
5	.375	.375	.5
6	.5	.5625	1
7	.75	.75	.5
8	.9375	1	1
9	1.25	1.25	.5
10	1.5	1.5625	1
11	1.875	1.875	.5
12	2.1875	2.25	1
13	2.625	2.625	.5
14	3	3.0625	1
15	3.5	3.5	.5
16	3.9375	4	1
17	4.5	4.5	.5
18	5	5.0625	1
19	5.625	5.625	.5
20	6.1875	6.25	1

As a check on our proof of the lemma, we note that when there is a difference in Table 5, the difference is $.0625 = 1/16$, the same as in Cases 1 and 3 of the proof of Lemma 1.

C. LOCAL OPTIMALITY OF THE DESIGNS:

Lemma. Let us choose designs for Langmuir's model. Given a sample, we add an $\mathbf{x}$ which is optimal. We set the first $\mathbf{x}$ equal to $\mathbf{x}_\infty$ without loss of generality. For even sample sizes, the result is half at $\frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty}$ and half

at $\mathbf{x}_\infty$. For odd sample sizes we take $(n \pm 1)/2$ observation at $\frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty}$ and $(n \mp 1)/2$ at $\mathbf{x}_\infty$.

Furthermore, this design cannot be improved on by replacing an observation with a best value conditioned on the remaining observations.

Proof of the Lemma

For ease, we work with ν, showing $\nu = \frac{1}{2}$ to yield $\mathbf{x} = \frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty}$, and $\nu = 1$ to yield $\mathbf{x} = \mathbf{x}_\infty$. From the previous argument that at least one ν_i must equal one, we set $\nu_1 = 1$. We add the nth observation.

Suppose the previous design had m of $\nu_i = \frac{1}{2}$, and $n - m - 1$ of $\nu = 1$. First we compute the $S(l,\ n)$'s. When is the smaller root $\frac{1}{2}$?

$$S(l,\ n) = m\left(\frac{1}{2}\right)^l + (n - m - 1)1^l = \frac{2^l n + (1 - 2^l)m - 2^l}{2^l}$$

$$S(2,\ n) = \frac{4n - 3m - 4}{4}$$

$$S(3,\ n) = \frac{8n - 7m - 8}{8}$$

$$S(4,\ n) = \frac{16n - 15m - 16}{16},$$

and

$$\nu_n = \frac{3S(3,\ n) - \left\{[3S(3,\ n)]^2 - 8S(2,\ n)S(4,\ n)\right\}^{\frac{1}{2}}}{4S(2,\ n)}.$$

The discriminant d is

$$d = \frac{64n^2 + 81m^2 + 64 - 144nm - 128n + 144m}{64}$$

$$= \left(\frac{8n - 9m - 8}{8}\right)^2.$$

This leads to

$$\nu_n = \frac{24n - 21m - 24 - |8n - 9m - 8|}{32n - 24m - 32}$$
$$= \begin{cases} \frac{1}{2} & \text{if } 8n - 9m - 8 \geq 0 \\ \frac{32n-30m-32}{32n-24m-32} & \text{otherwise.} \end{cases}$$

This means the potential relative maximum is at $\nu_n = \frac{1}{2}$ if eight or less $\nu_i = \frac{1}{2}$'s are present per nine ν_i's in the previous design.

Next, we compute and maximize G for m of $\nu = \frac{1}{2}$ and $n - m$ of $\nu = 1$.

$$G = \left(\sum_{j=1}^{n} \nu_j^2\right)\left(\sum_{j=1}^{n} \nu_j^4\right) - \left(\sum_{j=1}^{n} \nu_j^3\right)^2 = \frac{m(n-m)}{16},$$

and $\frac{\partial G}{\partial m} = \frac{n-2m}{16}$, so $m = \frac{n}{2}$ is ideal with $m = \frac{n\pm1}{2}$ being best if n is odd, since m must be an integer. Thus the first part of the Lemma generates a sequence of designs with $m = \frac{n}{2}$ for n even, and $m = \frac{n\pm1}{2}$ for n odd, satisfying the conditions for $\frac{1}{2}$ to be the second zero in each case. The second part of the Lemma similarly holds, the most extreme cases being even n after deletion and $m = \frac{n\pm2}{2}$, which again satisfies the condition $\frac{m}{n} \leq \frac{8}{9}$, using n instead of $n - 1$ since the n in our formula for G is $n - 1$ in the formula for the partial derivative. So the lemma is true.

IV. Simulation Based Results 2

We compare four designs. The first is $(1/n, \ldots n/n)$, called Arithmetic 1. The second is $(3/n, 6/n, \ldots 3n/n)$, called Arithmetic 3. The maximum value $\mathbf{x}_\infty = 3$ is chosen from talks with Anita Katti. The third design is our optimal allocation, with half the observations at $\mathbf{x}_\infty = 3$ and half at $\frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty} = .09375$. The fourth design, called Logarithmic, is uniform on the log scale with maximum value 3 and no replication.

Table 6	Means of $\hat{a}$, with $a = 25$.			
Sample Size:	10	40	160	640
Arithmetic 1:	26.81	25.37	25.13	25.03
Arithmetic 3:	36.17	25.67	25.21	25.04
Logarithmic:	33.47	26.80	25.49	25.12
Optimal:	25.41	25.14	25.05	25.00

The least value of $\mathbf{x}$ in the Logarithmic design is computed to minimize the generalized (asymptotic) variance, yielding $\mathbf{x}_1 = 0.327943$, 0.363068, 0.370807, and 0.372685 for sample sizes 10, 40, 160, and 640 respectively. The least value in the Logarithmic design is easy enough to compute since there is only one quantity to vary, namely the smallest $\mathbf{x}$ to be used.

Five samples out of 2000 in the Logarithmic design yielded poor results, namely least squares estimates of b in excess of 2000 with a true value of 10, and their estimates, $\hat{b}$, were set equal to zero. Similarly, five samples out of 2000 in the Arithmetic 3 design yielded poor results. Each of these five samples are included in the statistics for the logarithmic design and arithmetic designs respectively.

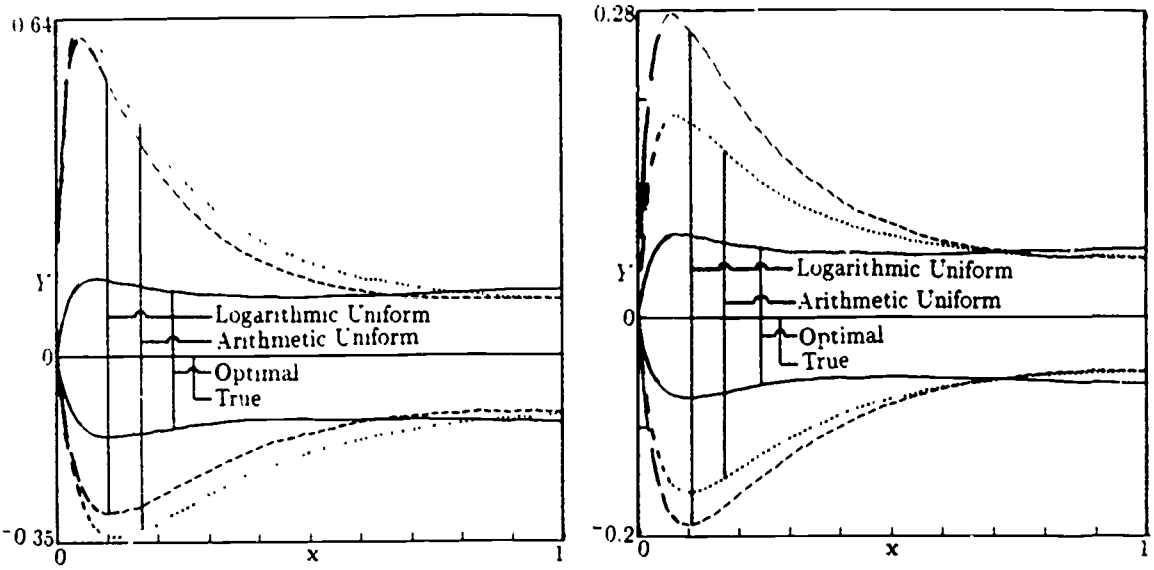
Note that the means for Arithmetic 1 and Optimal designs are fairly close to the true value of 25. For the (estimated) standard deviations, the **optimal design is the best, by a factor of $\sqrt{3}$ or better.**

Table 7	Standard Deviations of $\hat{a}$, with $a = 25$.			
Sample Size:	10	40	160	640
Arithmetic 1:	8.46	3.43	1.69	.86
Arithmetic 3:	98.35	4.71	2.19	1.11
Logarithmic:	47.22	8.32	3.46	1.68
Optimal:	3.99	1.97	.97	.47

Next, we look at graphs of the .90 midranges for the three designs. To make the comparisons easier, we subtract the true mean from the $\hat{Y}$'s. These midranges are obtained by computing at each value of $\mathbf{x}$, the 2000 values of $\hat{Y}$, rank them and graph the $\hat{Y} - \mathcal{E}(Y)$ at the .05 level, $\hat{y}_{.05} - \mathcal{E}(Y)$, and graph the $\hat{Y} - \mathcal{E}(Y)$ at the .95 level. The boundary curves of the midranges

cross each other repeatedly, so there is no great disadvantage in one design over another, aside from logarithmic uniform being a bit off for x around 0.2.

.90 Midranges from Different Designs



A. Sample Size n=10 B. Sample Size n=40

Fig. 3. Key: Arithmetic Uniform 3: short dashes; Optimal: thick solid; Logarithmic Uniform: long dashes.

As we increase the sample size, the midrange curves become considerably closer to the true curve and to each other.

We kept track of the ten samples from the two allocations which led to unacceptable estimates of b, that is estimates in excess of 2000 with a true value of 10. The samples are available on request. It appears that the logarithmic and "arithmetic 3" designs place so few observations near the origin and near the maximum value that the estimation is unstable.

In summary, there does not seem to be any serious loss associated with the use of our locally optimal design. To the contrary, **our design might reduce the sample size needed by a factor of three or better.**

V. Regarding Global Optimality

We check whether we can improve on our locally optimal solution by changing all of the observations simultaneously. Let $m = n/2$ or $m = (n \pm 1)/2$ as appropriate. We denote by d_i the changes to the first m values of ν_i and by $-f_i$, the changes to the second half. The ranges of d_i are $[-\frac{1}{2},\ \frac{1}{2}]$ and of f_i are $[0,\ 1]$. The expression for G becomes:

$$G = \left[\sum_{i=1}^{m}\left(\frac{1}{2}+d_i\right)^2 + \sum_{i=m+1}^{n}(1-f_i)^2\right]$$
$$\times \left[\sum_{i=1}^{m}\left(\frac{1}{2}+d_i\right)^4 + \sum_{i=m+1}^{n}(1-f_i)^4\right]$$
$$- \left[\sum_{i=1}^{m}\left(\frac{1}{2}+d_i\right)^3 + \sum_{i=m+1}^{n}(1-f_i)^3\right]^2$$

Next, we define

$$D_k = \sum_{i=1}^{m} d_i^k, \quad \text{and} \quad F_k = \sum_{i=m+1}^{n} f_i^k, \quad \text{for} \quad k = 1,\ 2,\ 3,\ 4.$$

We rewrite G as:

$$G = \left[\frac{4n-3m}{4} + D_1 + D_2 - 2F_1 + F_2\right]$$
$$\times \left[\frac{16n-15m}{16} + \frac{D_1}{2} + \frac{3D_2}{2} + 2D_3 + D_4 - 4F_1 + 6F_2 - 4F_3 + F_4\right]$$
$$- \left[\frac{8n-7m}{8} + \frac{3D_1}{4} + \frac{3D_2}{2} + D_3 - 3F_1 + 3F_2 - F_3\right]^2$$

The interesting part is that regardless of the value of n, the problem of optimal design reduces to these eight variables, $\mathbf{D} = (D_k;\ F_k)$. We expanded

the polynomial manually and by *Mathematica*. Since the polynomial is a quadratic form in eight variables, we write it in the matrix format as follows.

$$Q = \frac{m(n-m)}{16} + \mathbf{D}\begin{bmatrix} 0 \\ \frac{-8n+9m}{16} \\ \frac{m}{4} \\ \frac{4n-3m}{4} \\ \frac{-3m}{8} \\ \frac{16n-3m}{16} \\ \frac{-8n+5m}{4} \\ \frac{4m-3m}{4} \end{bmatrix}$$

$$+\mathbf{D}\begin{bmatrix} -\frac{1}{16} & -\frac{1}{8} & \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 1 & -\frac{5}{4} & \frac{1}{2} \\ -\frac{1}{8} & -\frac{3}{4} & -\frac{1}{2} & \frac{1}{2} & 1 & -\frac{3}{4} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & -1 & 0 & 1 & -2 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & -1 & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -\frac{3}{4} & -2 & \frac{1}{2} & 1 & -3 & 1 & \frac{1}{2} \\ -\frac{5}{4} & -\frac{1}{2} & 1 & 0 & 1 & 1 & -1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & -1 & \frac{1}{2} & 0 & 0 \end{bmatrix}\mathbf{D}^T$$

The eigenvalues **E**, of the 8×8 matrix and the matrix **M** of eigenvectors is:

$$\mathbf{E} = [.000 \quad .000 \quad .000 \quad .000 \quad .000 \quad 2.109 \quad 3.266 \quad 5.655]$$

$$\mathbf{M} = \begin{bmatrix} .160 & .654 & .421 & .044 & .066 & .473 & .324 & .185 \\ .589 & .051 & .012 & .121 & .663 & .046 & .382 & .222 \\ .344 & .312 & .245 & .402 & .489 & .324 & .088 & .460 \\ .435 & .466 & .566 & .177 & .170 & .369 & .247 & .124 \\ .024 & .288 & .383 & .027 & .434 & .461 & .516 & .320 \\ .477 & .129 & .333 & .131 & .270 & .114 & .121 & .726 \\ .253 & .373 & .428 & .239 & .034 & .414 & .582 & .212 \\ .168 & .125 & .036 & .846 & .164 & .369 & .247 & .124 \end{bmatrix}$$

It is interesting that five of the eight eigenvalues are zero and only three are nonzero. Thus, the optimal design does not depend on all eight variables independently. It depends on only three functions of them.

The range of the three functions is not nice. We could not proceed any further along this line. So, we checked on the global optimality by computing the function G over a fine grid, of size 100^8 over all of the eight variables. $\mathbf{D} = [D_k,\ F_k]$. With the above information, computation over such a large grid was not excessively time consuming. The computations confirmed that our design was globally optimal.

This is our current evidence that the design is globally optimal if half the observations are taken at the maximal value and half where the ν's equal half their maximal value.

VI. Conclusion

In conclusion, we suggest the use of whatever model is expected to have constant errors in the dependent variable. For the usual assumptions of normality and constant variance, we further suggest the design of half the observations to the maximum value, $\mathbf{x} = \mathbf{x}_\infty$, and half to $\mathbf{x} = \frac{\mathbf{x}_\infty}{2+b\mathbf{x}_\infty}$, to maximize the accuracy of parameter estimation. It is interesting that the statistical assumptions are as important as they seem. It is also interesting that the optimal design seems to be as simple and extreme as the optimal design in simple linear regression.

VII. Literature Cited

[1]D. Colquhoun, *Lectures on Biostatistics; an introduction to statistics with applications in biology and medicine* Clarendon, Oxford, 1971, pp. 257–272

[2]Irving Langmuir, (1916). "The constitution and fundamental properties of solids and liquids," *Journal of the American Chemical Society*. **38:** 2221

[3]Mohammed Habibullah, *Modification of Steepest Descent with Applications to Likelihood Functions*, Ph. D. dissertation, University of Missouri at Columbia, May 1988

[4]J. F. Eastham, V. N. LaRiccia, and J. H. Schuenemeyer, (1987). "Small sample properties of the maximum likelihood estimators for an alternative parameterization of the three-parameter lognormal distribution," *Commun. Statist. -Simula.*, **16**(3), 871–884.

[5]F. R. Oliver. (1970). "Some Asymptotic Properties of Colquhoun's Estimators for a Rectangular Hyperbola," *Applied Statistics,* or *Journal of the Royal Statistical Society (JRSS) series C*, **19:** 3 269–273.

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 21,1991